International Journal of Informatics and Computation (IJICOM)
Vol.7, No.1, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Sentiment Analysis of a 271 Trillion Rupiahs Corruption Case Using LSTM

Selamet Riadi[1], Rudi Muslim[2], Emi Suryadi[3], Karina Nurwijayanti[4], M. Zulpahmi[5], Muhamad Masjun Efendi[6], Bahtiar Imran[7]

Abstract

Corruption is one of the most pressing issues in Indonesia, significantly affecting public trust in governance and the nation's development. Among the many corruption cases that have surfaced, the recent 271 trillion rupiahs corruption case has drawn widespread attention and public discourse. Understanding the public's perception and sentiment regarding such cases can provide valuable insights into how these issues impact society. Researchers identified an opportunity to leverage sentiment analysis as a method to capture and analyze public sentiment in this context. The dataset for this study was collected from the social media platform Twitter (X) using a data crawling technique. Prior to analysis, preprocessing was performed to clean and prepare the data. After preprocessing, the data was categorized into three sentiment labels: negative, positive, and neutral. To perform sentiment classification, this study utilized the LSTM (Long Short-Term Memory) algorithm, a deep learning method particularly suited for sequential data analysis. The model was trained over a total of 10 epochs. The classification results demonstrated that the LSTM algorithm achieved an accuracy of 0.9365 at the 10th epoch, showcasing its effectiveness in analyzing public sentiment regarding 271 trillion rupiahs corruption issues.

Keywords:
Sentiment Analysis, Long Short-term Memory, Corruption

1. Introduction

Corruption practices still pose a major obstacle for many countries. particularly in Indonesia, significantly affecting governance, the economy, and public confidence. Corruption remains a parasite that disrupts the sustainability of democracy in Indonesia. A recent corruption case involving a state loss of 271 trillion Rupiah has sparked widespread public outrage, particularly concerning the relatively light sentence of 6.5 years imprisonment for the perpetrator. This disparity between the magnitude of the crime and the leniency of the sentence has ignited debates across social media such as Instagram, Twitter (X) and etc, providing an opportunity to analyze public sentiment on twitter about the issue, by utilizing sentiment analysis techniques using the LSTM algorithm [1].

The research team began conducting a literature review to support the references for the planned study. Subsequently, the researchers collected data from Twitter using data scraping techniques. Scraping is the process of retrieving data from a website [1].Followed by the text preprocessing stage. Preprocessing is the initial step in processing extracted text data before the data is processed further until it reaches the classification stage. The

Corresponding Author: Selamet Riadi, selametriadi@utmmataram.ac.id
1 Rudi Muslim, Universitas Teknologi Mataram, rudymuslim93@gmail.com
2 Emi Suryadi, Universitas Teknologi Mataram, emisuryadi@gmail.com
3. Karina Nurwijayanti, Universitas Teknologi Mataram, karinanurwijayanti.math@gmail.com
4. M. Zulpahmi, Universitas Teknologi Mataram, pahmijorge04@gmail.com
5. Muhamad Masjun Efendi, Universitas Teknologi Mataram, creativepio@gmail.com
6. Bahtiar Imran, Universitas Teknologi Mataram, bahtiarimranlombok@gmail.com

results of the text preprocessing were then used to classify tweet sentiments into positive, negative, or neutral categories with the help of an Indonesian sentiment lexicon. Next, the classification process was carried out based on the three predetermined sentiments. The classified data was then further analyzed [3].

The purpose of this study is to apply sentiment analysis techniques to classify public sentiment regarding corruption cases by utilizing text analysis methods such as preprocessing, word weighting, and deep learning algorithms, specifically Long Short-Term Memory (LSTM). Additionally, the study employs the confusion matrix evaluation technique to assess the classification results produced by the LSTM algorithm. This research is expected to serve as a reference for future studies on sentiment analysis. The findings from this study are expected to provide valuable insights into public sentiment regarding corruption cases in Indonesia, particularly in understanding how the public perceives discrepancies in sentencing for significant corruption offenses.

2. Related Works

There are several studies related to the research that the researchers plan to conduct computer and security [18]. A study was chosen LSTM as the method for sentiment classification. The dataset used is a movie review dataset consisting of 25,000 review documents, with an average length of 233 words per review. This study utilized the CBOW and Skip-Gram methods in Word2Vec to construct word vector representations for each word in the corpus. Several word vector dimensions, including 50, 60, 100, 150, 200, and 500, were applied to analyze their impact on the resulting accuracy. The results showed that the highest accuracy was achieved with a word vector dimension of 100 at 88.17%, while the lowest accuracy was 85.86% with a word vector dimension of 500 [4].

Another research conducts sentiment analysis on skincare product reviews using word embedding and the Long Short-Term Memory (LSTM) method. With a dataset of 1,500 entries and utilizing Word2Vec word embedding, the study achieved an accuracy of 91% for 90% training data and 10% testing data. The accuracy decreased to 87% for 80% training data and 20% testing data, and 85% for 70% training data and 30% testing data. Without Word2Vec, the highest accuracy was only 74% for 90% training data and 10% testing data, with lower accuracy observed for other proportions [5]. A study conducts sentiment analysis by combining deep learning algorithms, specifically Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). Word2Vec is also utilized for word weighting. The employed model successfully achieved a highest accuracy of 85% [6]. There is research also using LSTM, like research from [7] that conducted research to predict stock prices.

A paper explored research using the Multinomial Naive Bayes algorithm to classify public sentiment toward the government's handling of COVID-19. The data used in the research was collected from Twitter (X), with TF-IDF employed as the word weighting technique and the confusion matrix used as the evaluation model. The built model achieved an accuracy of 74%, which is a fairly good result considering the model and the dataset consisting of 2,000 entries [8]. Another article conducted sentiment analysis research on cyberbullying by comparing the Multinomial Naive Bayes (MNB) algorithm and the Support Vector Machine (SVM) algorithm. The study also incorporated a feature selection technique, namely chi-square, to enhance the accuracy of the developed models. The results showed that the MNB algorithm achieved an accuracy of 90.77% when applying chi-square, compared to only 83.85% without chi-square. Meanwhile, the SVM algorithm achieved an

accuracy of 90.00% with chi-square, whereas without chi-square, it only reached an accuracy of 82.31% [9]. Current research conducted sentiment analysis on user reviews of the Webtoon application using the Support Vector Machine (SVM) classification algorithm. The accuracy achieved by the developed model was 82%. It highlights the need to explore other classification algorithms, such as deep learning [10].

3. Proposed Method

To achieve the research objectives, our study conducted several steps including a literature review to support the research we planned to undertake. Next, we collected data from Twitter using a data crawler technique. Then, we performed preprocessing on the crawled data. Following that, we applied word weighting and divided the data into training and testing sets, which were then directly input into the algorithm—specifically, the LSTM algorithm used in this study. Fig. 1 depict the model framework of this study.

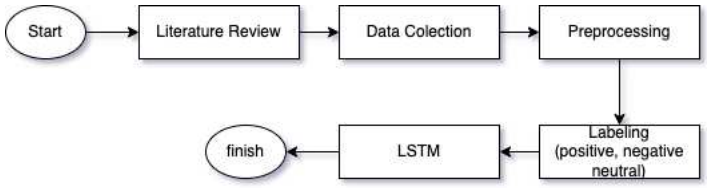

Fig.1 Model Framework using LSTM

The research framework in Fig.1 will be explained through the following points:

3.1. Start

The research framework illustrated begins with the Start phase, which marks the initiation of the study.

3.2. Literature Review

In this step, the Data Collection phase is conducted to gather the necessary data for the research. This data may include textual information, social media posts, or other relevant datasets that align with the study's objectives.

3.3. Data Collection

This research utilizes tweets collected from the social media platform Twitter (X). A crawling technique was employed to gather data from Twitter (X). The data was collected between January 1, 2024, and December 29, 2024, resulting in a total of 1,005 rows of data.

3.4. Preprocessing

Once the data is collected, it undergoes a Preprocessing stage. During this stage, the raw data is cleaning, case folding, tokenization, stop words removal, normalization, and stemming.

1. This study use cleaning data technique. Data cleaning is a crucial preprocessing step to eliminate unnecessary noise before conducting further analysis [11].
2. Case-folding is a process of converting text documents into lowercase form [12]. Converting data with uppercase letters to lowercase is necessary to reduce irrelevant differences when comparing text, allowing data analysis to be conducted more consistently and accurately.

3. Tokenization is a process of selecting and trimming words in a sentence [13]. This process is necessary to simplify text, focus on relevant information, and improve the efficiency of data analysis.
4. Stop words are word features that do not contain any sentiment elements, such as conjunctions like 'moreover,' 'or,' 'from,' 'but,' 'and,' and so on [14]. This process is necessary to remove irrelevant words, reduce noise in the data, and enhance the accuracy of text analysis.
5. Text normalization is the process of transforming written text into a format that reflects how it should be spoken [15]. This process is crucial for ensuring consistency, improving the readability of text, and enabling accurate analysis or text-to-speech conversion.
6. Stemming is a step-in text processing that focuses on identifying the root form of a word found in a text [16]. This process aims to reduce words to their base form, enabling more efficient text analysis and improving the consistency of data by grouping similar words together.

3.5. Labeling

A critical component of this framework is the Labeling process, where the data is categorized into predefined sentiment classes: positive, negative, or neutral. This labeled dataset is crucial for training the model.

3.6. LSTM

LSTM is an advancement of the Recurrent Neural Network (RNN) algorithm designed to address the challenges RNN faces in handling long-term data [17]. The LSTM model operates based on the following key components and formulas:

1. **Forget Gate:** Determines which information from the previous state should be forgotten. Table 1 summarizes the mathematical of **forget gate** in LSTM.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f \quad (1)$$

Table 1. The Mathematical Notation of Forget Gate

Notation	Description
f_i	$forget\ gate\ output$
σ	$Sigmoid\ activation\ fuction$
W_f	$Weights\ for\ the\ forget\ gate$
h_{t-1}	$Previous\ hidden\ state$
x_i	$input\ at\ the\ current\ timestep$
b_f	$Bias\ for\ the\ forget\ gate$

2. **Input Gate:** Decides which new information to store in the cell state.

$$\begin{gathered} i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c.[h_{t-1}, x_t] + b_c) \quad (2) \\ C_t = f_t.C_{t-1} + i_t.\sim C_t \end{gathered}$$

Table 2. The Mathematical Notation of Input Gate

Notation	Description

i_t	$input\ gate\ output$
$\tilde{C}_t$	$Candidate\ cell\ state$
C_t	$Update\ cell\ state$
W_i, W_c	$Weights\ for\ input\ gate\ and\ candidate\ state$
b_i, b_c	$Bias\ terms$

3. **Output Gate:** Determines the output of the LSTM cell.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (3)$$
$$h_t = o_t.\tanh(C_t)$$

Table 3. The Mathematical Notation of Output gate

Notation	Description
o_t	$Output\ gate\ output$
h_t	$Hidden\ state\ (output\ of\ the\ LSTM)$
W_o	$Weights\ for\ the\ output\ gate$
b_o	$Bias\ for\ the\ output\ gate$

These mechanisms collectively allow the LSTM to maintain relevant information over extended sequences, effectively capturing dependencies that span across multiple timesteps. The LSTM processes the labeled data and outputs predictions for sentiment classification, such as positive, negative, or neutral sentiments. Finally, the study concludes at the end phase, where insights from the analysis are summarized, and the research objectives are addressed.

4. Experimental Setup

The experimental setup for this study involves several key stages, each contributing to the process of sentiment analysis using the LSTM model:

1. Literature Review
   The experiment begins with an extensive literature review. This step is crucial to understand the existing methodologies, tools, and techniques in sentiment analysis and deep learning, particularly focusing on the use of LSTM models. Insights gained from the review help in defining the research objectives and designing the experiment.
2. collecting data
   The next step involves collecting data relevant to the sentiment analysis task. The dataset in this research was obtained from the social media platform Twitter (X) using the keyword 'korupsi 271' through data crawling techniques. The top 5 raw datasets in this study can be seen in Table 4.

Table 4. 5 Raw Dataset

No	Created_at	Full_text	location	Tweet_url	User_id_str	username
1	Mon Dec 30 21:51:41 +0000 2024	Hukuman maksimal dijatuhkan kepada pelanggar lalu lintas karena SIM sudah tidak berlaku & penumpang dibelakang tidak memakai helm sebesar Rp. 1.250.000. Namun hukuman ringan dijatuhkan kepada koruptor karena korupsi sebesar Rp. 271 T	Indonesia	https://x.com/riswan_suhu/status/1873849462977974478	79115455	riswan_suhu

		berupa kurungan penjara 6 Tahun 6 bulan.				
2	Mon Dec 30 16:22:45 +0000 2024	*Netizen: Hukuman Sosial Akan Diterima Bagi Orang Ini Dan Keluarganya.* jaksa yg mengadili Harvey Muis dgn Vonis 6 5 th Korupsi 271 Triliyuncukup janggal keputusan tsb . □üáÆ□üá©ÔÎè https://t.co/pdQ0r1CFvm	Jabodetabek, Indonesia	https://x.com/Andria75777/status/1873766684726816940	1.6672E+18	Andria75777
3	Mon Dec 30 16:22:45 +0000 2024	*Netizen: Hukuman Sosial Akan Diterima Bagi Orang Ini Dan Keluarganya.* jaksa yg mengadili Harvey Muis dgn Vonis 6 5 th Korupsi 271 Triliyuncukup janggal keputusan tsb . □üáÆ□üá©ÔÎè https://t.co/pdQ0r1CFvm	Jabodetabek, Indonesia	https://x.com/Andria75777/status/1873766684726816940	1.66723E+18	Andria75777
4	Mon Dec 30 16:02:06 +0000 2024	@cobeh2022 horeeee ... ini loh hasil korupsi 271 T jalan2 ke eropa klo sakit kan ada BPJS PBI	rumah	https://x.com/walangfighter/status/1873761488101617916	70867143	walangfighter
5	Mon Dec 30 15:59:23 +0000 2024	Korupsi Timah; Hakim tetapkan Kerugian Lingkungan Hidup senilai Rp.271 triliun ÔÎè Hukuman penjara 6 5 tahun atau selama 2 372.5 hari ÔÎè Keuntungan Si Koruptor setiap detiknya sebesar Rp1 2 juta https://t.co/Nm0BT1FJN4	Indonesia	https://x.com/roysalam/status/1873760803784134698	80254922	roysalam

3. Preprocessing

After data collection, preprocessing is performed to clean and prepare the raw data. This step includes cleaning, case folding, tokenization, stop words removal, normalization, and stemming. Preprocessing ensures that the data is suitable for model training and reduces noise. The top 5 result of preprocessing data we can si in Table 5.

Table 5. Top 5 Result of Preprocessing

no	full_text	cleaned_text	case_folded_text	tokenized_text	filtered_text	text_normalize	stemming_data
1	Hukuman maksimal dijatuhkan kepada pelanggar lalu lintas karena SIM sudah tidak berlaku & penumpang dibelakang tidak memakai helm sebesar Rp. 1.250.000. Namun hukuman ringan dijatuhkan kepada koruptor karena korupsi sebesar Rp. 271 T berupa kurungan penjara 6 Tahun 6 bulan.	hukuman maksimal dijatuhkan kepada pelanggar lalu lintas karena sim sudah tidak berlaku amp penumpang dibelakang tidak memakai helm sebesar rp namun hukuman ringan dijatuhkan kepada koruptor karena korupsi sebesar rp t berupa kurungan penjara tahun bulan	hukuman maksimal dijatuhkan kepada pelanggar lalu lintas karena sim sudah tidak berlaku amp penumpang dibelakang tidak memakai helm sebesar rp namun hukuman ringan dijatuhkan kepada koruptor karena korupsi sebesar rp t berupa kurungan penjara tahun bulan	['hukuman', 'maksimal', 'dijatuhkan', 'kepada', 'pelanggar', 'lalu', 'lintas', 'karena', 'sim', 'sudah', 'tidak', 'berlaku', 'amp', 'penumpang', 'dibelakang', 'tidak', 'memakai', 'helm', 'sebesar', 'rp', 'namun', 'hukuman', 'ringan', 'dijatuhkan', 'kepada', 'koruptor', 'karena', 'korupsi', 'sebesar', 'rp', 't', 'berupa', 'kurungan', 'penjara', 'tahun', 'bulan']	['hukuman', 'maksimal', 'dijatuhkan', 'pelanggar', 'lintas', 'sim', 'berlaku', 'amp', 'penumpang', 'dibelakang', 'memakai', 'helm', 'rp', 'hukuman', 'ringan', 'dijatuhkan', 'koruptor', 'korupsi', 'rp', 't', 'kurungan', 'penjara']	['hukuman', 'maksimal', 'dijatuhkan', 'pelanggar', 'lintas', 'sim', 'berlaku', 'amp', 'penumpang', 'dibelakang', 'memakai', 'helm', 'rp', 'hukuman', 'ringan', 'dijatuhkan', 'koruptor', 'korupsi', 'rp', 't', 'kurungan', 'penjara']	hukum maksimal jatuh langgar lintas sim laku amp tumpang belakang pakai helm rp hukum ringan jatuh koruptor korupsi rp t kurung penjara
2	*Netizen: Hukuman Sosial Akan Diterima Bagi Orang Ini Dan Keluarganya.* jaksa yg mengadili Harvey Muis dgn Vonis 6 5 th Korupsi 271 Triliyuncukup janggal keputusan tsb . □üáÆ□üá©ÔÎè https://t.co/pdQ0r1CFvm	netizen hukuman sosial akan diterima bagi orang ini dan keluarganya jaksa yg mengadili harvey muis dgn vonis th korupsi triliyuncukup janggal keputusan tsb	netizen hukuman sosial akan diterima bagi orang ini dan keluarganya jaksa yg mengadili harvey muis dgn vonis th korupsi triliyuncukup janggal keputusan tsb	['netizen', 'hukuman', 'sosial', 'akan', 'diterima', 'bagi', 'orang', 'ini', 'dan', 'keluarganya', 'jaksa', 'yg', 'mengadili', 'harvey', 'muis', 'dgn', 'vonis', 'th', 'korupsi', 'triliyuncukup', 'janggal', 'keputusan', 'tsb']	['netizen', 'hukuman', 'sosial', 'diterima', 'orang', 'keluarganya', 'jaksa', 'yg', 'mengadili', 'harvey', 'muis', 'dgn', 'vonis', 'th', 'korupsi', 'triliyuncukup', 'janggal', 'keputusan', 'tsb']	['netizen', 'hukuman', 'sosial', 'diterima', 'orang', 'keluarganya', 'jaksa', 'yg', 'mengadili', 'harvey', 'muis', 'dgn', 'vonis', 'th', 'korupsi', 'triliyuncukup', 'janggal', 'keputusan', 'tsb']	netizen hukum sosial terima orang keluarga jaksa yg adil harvey mu dgn vonis th korupsi triliyuncukup janggal putus tsb
3	@cobeh2022 horeeee ... ini loh hasil korupsi 271 T jalan2 ke eropa klo sakit kan ada BPJS PBI	cobeh horeeee ini loh hasil korupsi t jalan ke eropa klo sakit kan ada bpjs pbi	cobeh horeeee ini loh hasil korupsi t jalan ke eropa klo sakit kan ada bpjs pbi	['cobeh', 'horeeee', 'ini', 'loh', 'hasil', 'korupsi', 't', 'jalan', 'ke', 'eropa', 'klo', 'sakit', 'kan', 'ada', 'bpjs', 'pbi']	['cobeh', 'horeeee', 'loh', 'hasil', 'korupsi', 't', 'jalan', 'eropa', 'klo', 'sakit', 'bpjs', 'pbi']	['cobeh', 'horeeee', 'loh', 'hasil', 'korupsi', 't', 'jalan', 'eropa', 'klo', 'sakit', 'bpjs', 'pbi']	cobeh horeeee loh hasil korupsi t jalan eropa klo sakit bpjs pbi
4	Korupsi Timah; Hakim tetapkan Kerugian Lingkungan Hidup senilai Rp.271 triliun ÔÎè Hukuman penjara 6 5 tahun atau selama 2 372.5 hari ÔÎè	korupsi timah hakim tetapkan kerugian lingkungan hidup senilai rp triliun hukuman penjara tahun atau selama hari keuntungan si koruptor setiap	korupsi timah hakim tetapkan kerugian lingkungan hidup senilai rp triliun hukuman penjara tahun atau selama hari keuntungan si koruptor setiap	['korupsi', 'timah', 'hakim', 'tetapkan', 'kerugian', 'lingkungan', 'hidup', 'senilai', 'rp', 'triliun', 'hukuman', 'penjara', 'tahun', 'atau', 'selama', 'hari', 'keuntungan', 'si', 'koruptor', 'setiap',	['korupsi', 'timah', 'hakim', 'tetapkan', 'kerugian', 'lingkungan', 'hidup', 'senilai', 'rp', 'triliun', 'hukuman', 'penjara', 'keuntungan', 'si', 'koruptor', 'detiknya', 'rp', 'juta']	['korupsi', 'timah', 'hakim', 'tetapkan', 'kerugian', 'lingkungan', 'hidup', 'senilai', 'rp', 'triliun', 'hukuman', 'penjara', 'keuntungan', 'si', 'koruptor', 'detiknya', 'rp', 'juta']	korupsi timah hakim tetap rugi lingkung hidup nila rp triliun hukum penjara untung si koruptor detik rp juta

	Keuntungan Si Koruptor setiap detiknya sebesar Rp1 2 juta https://t.co/Nm0BT1FJN4	detiknya sebesar rp juta	detiknya sebesar rp juta	'detiknya', 'sebesar', 'rp', 'juta']			
5	@NataliusPigai2 Pak pinggai p√†k pinggai..... tidakah bapak berpikir bahwa korupsi itu melanggar HAM kalau lah uang yg di korup sebesar 271 T Dibagikan pada rakyat indo kira dapat berapa per kepala ..... Ah bapak makin ga jelas	nataliuspigai pak pinggai p√†k pinggai tidakah bapak berpikir bahwa korupsi itu melanggar ham kalau lah uang yg di korup sebesar t dibagikan pada rakyat indo kira dapat berapa per kepala ah bapak makin ga jelas	nataliuspigai pak pinggai p√†k pinggai tidakah bapak berpikir bahwa korupsi itu melanggar ham kalau lah uang yg di korup sebesar t dibagikan pada rakyat indo kira dapat berapa per kepala ah bapak makin ga jelas	['nataliuspigai', 'pak', 'pinggai', 'p√†k', 'pinggai', 'tidakah', 'bapak', 'berpikir', 'bahwa', 'korupsi', 'itu', 'melanggar', 'ham', 'kalau', 'lah', 'uang', 'yg', 'di', 'korup', 'sebesar', 't', 'dibagikan', 'pada', 'rakyat', 'indo', 'kira', 'dapat', 'berapa', 'per', 'kepala', 'ah', 'bapak', 'makin', 'ga', 'jelas']	['nataliuspigai', 'pinggai', 'p√†k', 'pinggai', 'tidakah', 'berpikir', 'korupsi', 'melanggar', 'ham', 'uang', 'yg', 'korup', 't', 'dibagikan', 'rakyat', 'indo', 'kepala', 'ah', 'ga']	['nataliuspigai', 'pinggai', 'p√†k', 'pinggai', 'tidakah', 'berpikir', 'korupsi', 'melanggar', 'ham', 'uang', 'yg', 'korup', 't', 'dibagikan', 'rakyat', 'indo', 'kepala', 'ah', 'ga']	nataliuspigai pinggai p k pinggai tidakah pikir korupsi langgar ham uang yg korup t bagi rakyat indo kepala ah ga

Prior to analysis, duplicate data was removed, reducing the dataset from 1,005 rows to 994 rows.

4. Labeling

   The cleaned data is then labeled into predefined sentiment categories: positive, negative, or neutral. This step is crucial for supervised learning, as it provides the ground truth for training the LSTM model. The result of labeling data we can see in the Fig. 2.

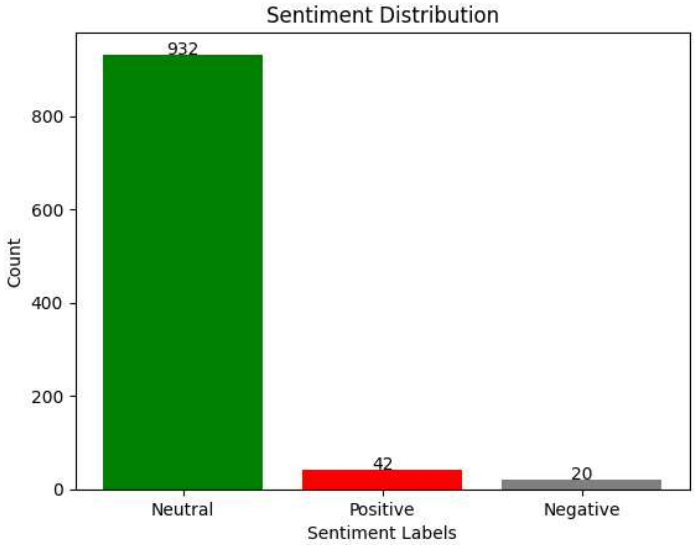

Fig. 2 Data Labeling

As we observe, the data contains overfitting, which naturally affects the accuracy results. However, since the objective of this study is to understand public sentiment regarding the 271 trillion corruption case, we proceeded to the classification process without applying any data balancing techniques

5. LSTM Classification

   In this study, classification was conducted using Long Short-Term Memory (LSTM), a deep neural network architecture highly effective for processing sequential data such as text. The LSTM model is designed to capture patterns and dependencies in long sequences, making it suitable for sentiment analysis.

   1.1. Model Architecture.

   The LSTM model was built with the following steps:

   1.1.1. **Embedding Layer**: This layer is used to represent words in numerical vector form, enabling the model to capture the semantic relationships

between words.

1.1.2. **LSTM Layer**: An LSTM layer with 100 units was employed to learn the sequential patterns of the text data. A dropout of 0.2 was applied to prevent overfitting.

1.1.3. **Fully Connected Layer**: A dense layer with 64 neurons and a ReLU activation function was added, followed by a dropout of 0.2 to enhance the model's generalization capabilities.

1.1.4. **Output Layer**: The output layer utilized the softmax activation function to produce classification probabilities for each target category.

5. Result and Analysis

According to the experiments result, the LSTM model achieved the highest accuracy of 0.9365, making it the best-performing model compared to other approaches in this study. These results demonstrate that LSTM effectively captures patterns and relationships in text data, making it highly suitable for sentiment analysis tasks. To see the accuracy result can be seen in the Fig. 3 and Table 6.

.

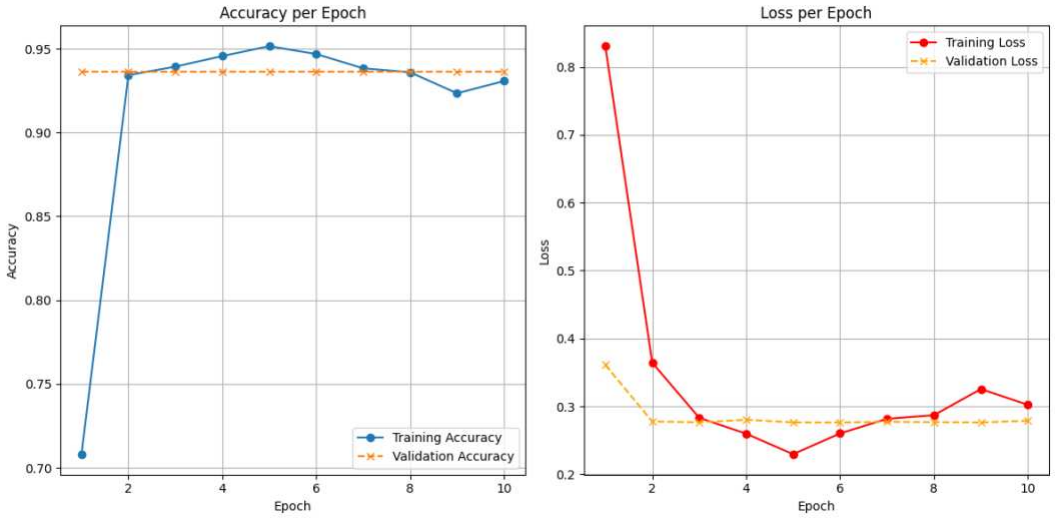

Fig. 3 Classification Result

Table 6: The proposed model accuracy in tabular form.

Epoch	Accuracy	Loss	Val accuracy	Val. loss
1	0.7081	0.8306	0.9365	0.3610
2	0.9342	0.3642	0.9365	0.2777
3	0.9393	0.2831	0.9365	0.2765
4	0.9456	0.2597	0.9365	0.2802
5	0.9515	0.2295	0.9365	0.2764
6	0.9468	0.2601	0.9365	0.2762
7	0.9383	0.2816	0.9365	0.2773
8	0.9356	0.2870	0.9365	0.2767

9	0.9234	0.3255	0.9365	0.2763
10	0.9307	0.3022	0.9365	0.2788

The accuracy achieved in this study reached 0.9365 at the 10th epoch, indicating that the performance of the developed model is highly satisfactory. However, it is important to note that the dataset used in this study exhibits overfitting, which may affect the overall performance and generalizability of the model. Additionally, the dataset consisted of only 1,000 samples, which is considered relatively small. Despite these limitations, the classification process was conducted without applying data balancing techniques, as the primary objective of this study was to analyze public sentiment regarding the specific topic.

6. Conclusion

This study successfully achieved its objective of analyzing public sentiment regarding the “271 trillion corruption” case using the Long Short-Term Memory (LSTM) algorithm, achieving an accuracy of 0.9365. The findings demonstrate the capability of LSTM to process complex linguistic patterns and provide meaningful insights, highlighting its significant potential for effectively analyzing social issues. However, the study faced several limitations, including the use of a small and imbalanced dataset, which affected model accuracy and led to overfitting. The preprocessing and feature extraction methods also require improvements to better handle noise and capture nuanced linguistic features. These limitations emphasize the importance of high-quality data in machine learning studies to ensure more reliable outcomes.

Future research is encouraged to utilize larger and more diverse datasets to enhance the model's generalizability and reduce the risk of overfitting. Additionally, the development of advanced preprocessing strategies and the exploration of hybrid models could further refine algorithm performance. Incorporating domain-specific knowledge and external resources would also enrich the analysis and support more accurate sentiment classification.

Acknowledgment

I appreciate my team for their excellent cooperation, and I hope we will continue to have extraordinary collaboration in the future.

References

[1] F. Pahlevi, "Eradication of Corruption in Indonesia from the Perspective of Lawrence M. Friedman's Legal System," *El-Dusturie*, vol. 1, no. 1, 2022.

[2] M. I. Ghozali, W. H. Sugiharto, and A. F. Iskandar, "Sentiment Analysis of Online Loans on Twitter Social Media Using the Naive Bayes Method," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, no. 6, pp. 1340–1348, 2023.

[3] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Application of the SVM Algorithm for Sentiment Analysis on Twitter Data of the Corruption Eradication Commission of the Republic of Indonesia," *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, vol. 7, no. 1, pp. 1–11, 2020.

[4] W. Widayat, "Movie Review Sentiment Analysis Using Word2Vec and LSTM Deep Learning Methods," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 1018, Jul. 2021, doi: 10.30865/mib.v5i3.3111.

[5] R. G. Purnasiwi, K. Kusrini, and M. Hanafi, "Sentiment Analysis on Skincare Product Reviews Using Word Embedding and Long Short-Term Memory (LSTM) Method," *Innovative: Journal of Social Science Research*, vol. 3, no. 2, pp. 11433–11448, 2023.

[6] Z. B. Nezhad and M. A. Deihimi, "A Combined Deep Learning Model for Persian Sentiment Analysis," *IIUM Engineering Journal*, vol. 20, no. 1, pp. 129–139, 2019, doi: 10.31436/iiumej.v20i1.1036.
[7] H. Hamzah, "Stock Price Prediction in Indonesia's Mining Sector Using a Hybrid Conv1D-LSTM Model," International Journal of Informatics and Computation, vol. 6, no. 1, pp. 33–39, 2024.
[8] N. Hidayah and S. Sahibu, "Multinomial Naïve Bayes Algorithm for Sentiment Classification of Government Response to COVID-19 Using Twitter Data," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, 2021.
[9] S. Riadi, E. Utami, and A. Yaqin, "Comparison of NB and SVM in Sentiment Analysis of Cyberbullying Using Feature Selection," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 7, no. 4, pp. 2414–2424, 2023.
[10] A. Fergina, I. L. Kharisma, and A. I. Taek, "Sentiment Analysis of Webtoon Application User Reviews Using Text Mining and Algorithms," *Jurnal Informasi dan Teknologi*, pp. 164–170, 2023.
[11] S. Liu *et al.*, "Investigating Data Cleaning Methods to Improve Performance of Brain–Computer Interfaces Based on Stereo-Electroencephalography," *Frontiers in Neuroscience*, vol. 15, p. 725384, 2021.
[12] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, "Sentiment Analysis of Nusantara Capital Relocation Using Naïve Bayes and KNN Algorithms," *Jurnal KomtekInfo*, pp. 1–7, 2023.
[13] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Public Sentiment Analysis of Jakarta Government's Lockdown Policy Using the SVM Algorithm," *Jurnal Data Mining dan Sistem Informasi*, vol. 2, no. 1, pp. 31–37, 2021.
[14] S. Khomsah and A. S. Aribowo, "Text-Preprocessing Model for Indonesian YouTube Comments," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 4, pp. 648–654, 2020.
[15] A. H. Ng, K. Gorman, and R. Sproat, "Minimally Supervised Written-to-Spoken Text Normalization," in *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, IEEE, 2017, pp. 665–670.
[16] W. Rifai and E. Winarko, "Modification of Stemming Algorithm Using a Non-Deterministic Approach for Indonesian Text," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 4, pp. 379–388, 2019.
[17] M. F. Rizkilloh and S. Widiyanesti, "Cryptocurrency Price Prediction Using Long Short-Term Memory (LSTM) Algorithm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 25–31, 2022.
[18] P. Wanda, "A Survey of Intrusion Detection System," *International Journal of Informatics and Computation*, vol. 1, no. 1, pp. 1-10, Jan. 2020.